CS783 Assignment-4

Amrit Singhal (150092)
Department of Computer Science
IIT Kanpur
Kanpur, India
amrits@iitk.ac.in

Aarsh Prakash Agarwal (150004)
Department of Electrical Engineering
IIT Kanpur
Kanpur, India
aarshp@iitk.ac.in

1 Problem Statement

The objective of this assignment was Image segmentation in an unsupervised setting. Image segmentation is basically assigning every pixel value a label so that the same labels share similar properties. It can be of two types: semantic and instance segmentation. In instance segmentation, each instance of class of objects is assigned different label while in semantic segmentation, each class of objects is assigned with different label. For example, suppose you have an image with several entities such as trees, persons, cars. Assigning each person, car and tree a different label would be instance segmentation while assigning just three labels, each for person, car and tree is semantic segmentation. Our goal is to perform semantic image segmentation.

Dataset that was provided for this comprised of five camera angles of road in IIT Kanpur. Each camera had about 1000 continuous snapshots of mainly people, scooters, cars and road with a divider. Since no labels or annotations were provided for this task, entire project had to be carried out in completely unsupervised manner. We started initially with some low level image processing techniques such as thresholding and clustering and then moved on to some deep learning based approaches such as w-net. All the approaches with some of the results are discussed below.

2 Image processing based experiments

This section summarizes the experiments and results based on the low level processing of the images. We have various methods such as thresholding and K-means clustering [[2]].

2.1 Thresholding

Thresholding is a very simple technique and generally used for foreground vs background with single threshold values. It basically assigns each pixel value to a certain label if lies in a certain range. The method is pretty basic and therefore did not do well in the complex setting of street view. Below is a sample from dataset.

Figure 1: Thresholding Result

As we can see from Figure 1, only leaves in the trees seemed to have been segmented correctly. The roof of the car have the same shade as that of the road with bright sunlight area. Most of segmentation is based on the brightness of different regions. Tying on further we tried K-Means clustering.

2.2 K- Means clustering

As the name suggests we cluster the different regions image based on their pixel values. Here, we have decided to go ahead with the K-means so that the regions which are not connected can also be clustered.

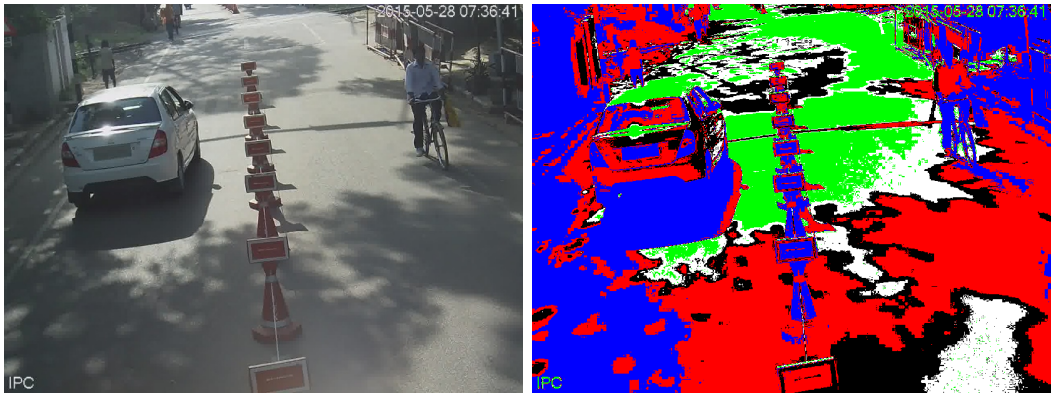

Figure 2: Clustering Result

We tried with different values of K and best result we got was for K =5. As we can see from the Figure 2, the shadow of trees on the road are clustered in red while the shadow of car and some other trees are clustered in blue while the brighter side of road is clustered in green colour which shows the inconsistency in this method. In the next section we will discuss the deep learning based methods.

3 Deep Learning Based methods

Next, we tried using certain methods that utilize the strength of deep learning to perform unsupervised image segmentation. We tried two such methods, namely image segmentation using backpropagation [1] and W-Net [3].

3.1 Using Back-propagation

This method is again based on the clustering the different sections of the image. Since the setting is unsupervised with no labels, it jointly optimizes the features obtained from a convolutional net and pixel labels while their parameters are being updated in an alternative fashion to promote assigning the pixels with similar features, similar label. It also promotes the pixels which are closer being assigned same label. It's architecture can be seen in the Figure 3

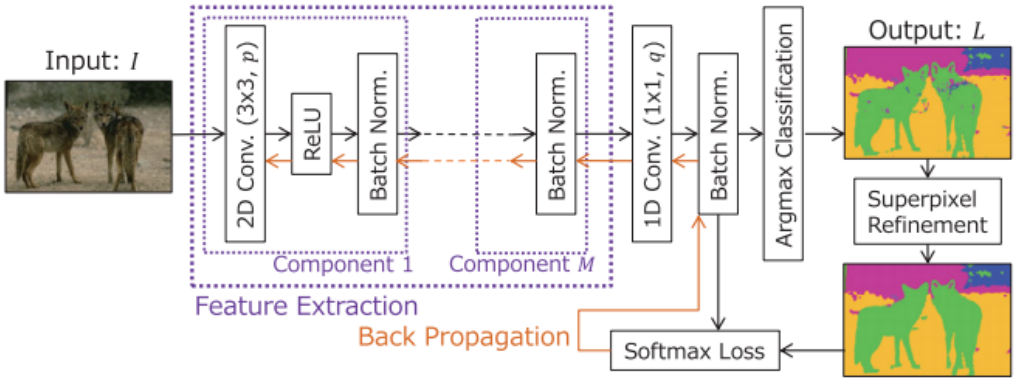

Figure 3: Back propagation method

The training with max-iterations around 1000 was done on the dataset. The decrease in the overall loss values while training can be seen from Figure 4

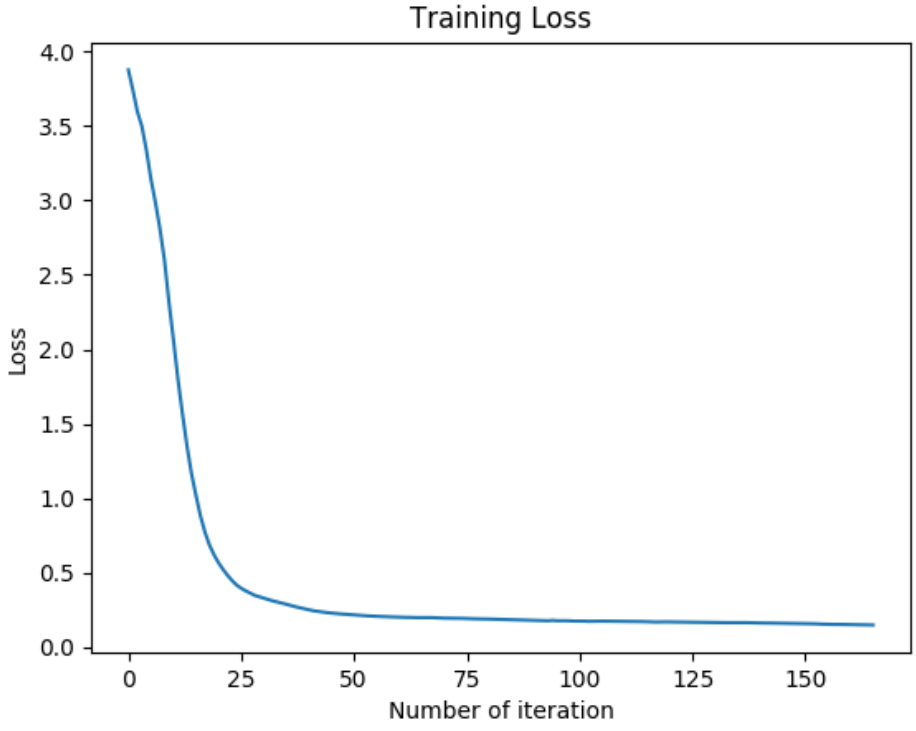

Figure 4: Training Loss

This method seemed to produce the best results which is because of it's efficiency to balance between assigning closer pixels same label, and pixels with same features same label. For example we can see in Figure 5, the shadows along with th vehicles are perfectly represented in the segmented image. Also, all the orange dividers on the road are assigned the same label.

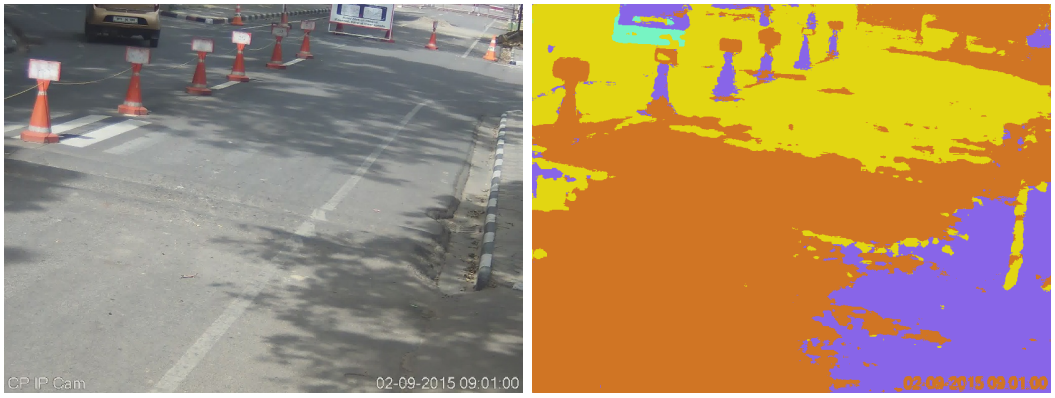

Figure 5: Backpropagation Result 1

3.2 W-Net

W-Net is the state-of-the-art deep learning architecture for completely unsupervised image semantic segmentation. The architecture consists of two U-nets are as an autoencoder to generate semantic segmentation of images. The authors propose using a quantity known as the Normalized cut, along with the autoencoder reconstruction loss, to update the parameters using backpropagation. The architecture of the W-net can be seen in Figure 6.

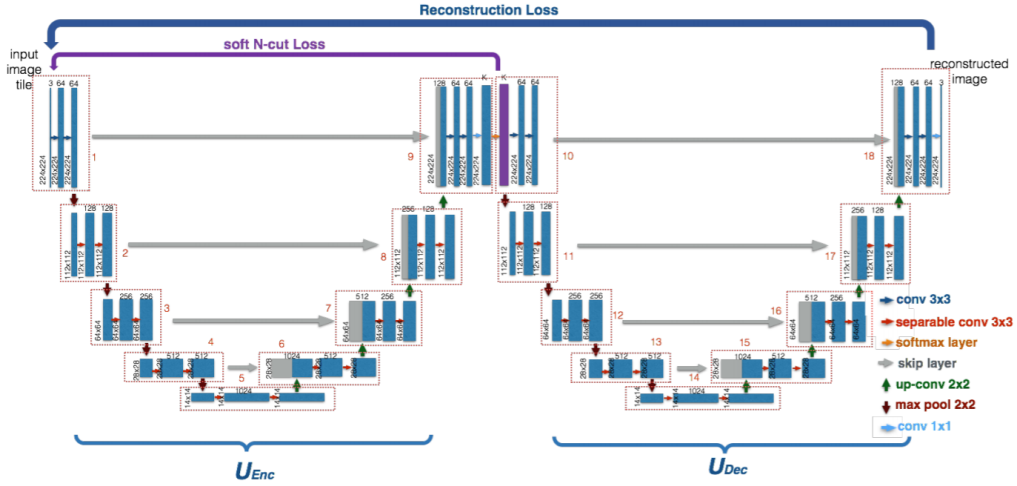

Figure 6: W-Net architecture

The model was trained for 50000 iterations, with a batch size of 5. The output from the model on certain sample images in shown in figure 7.

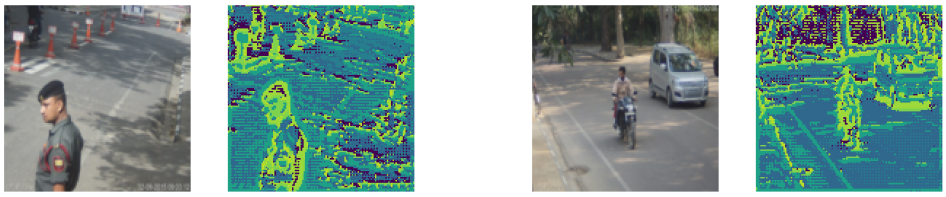

Figure 7: W-net sample results

We can observe that the results seen are much like those presented by the authors in the paper. The authors have also used a post-processing step of applying a CRF smoothing to get better looking segmentation. We have not done this step, but the results obtained till before that step look promising. The image looks to have been segmented properly into sections, divided on the edges. Clearly, these results are better than any other obtained previously.

References

[1] Asako Kanezaki. Unsupervised image segmentation by backpropagation. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1543–1547, 2018.

[2] Pulkit Sharma. Tutorial on image segmentation. `https://www.analyticsvidhya.com/blog/2019/04/introduction-image-segmentation-techniques-python/?fbclid=IwAR16VFtUo7oAK5MjyQWnhEuLbU0e5vlrQT29a-G624rfGtv952tDU9I-2ks`.

[3] Xide Xia and Brian Kulis. W-net: A deep model for fully unsupervised image segmentation. *CoRR*, abs/1711.08506, 2017.